

**USE OF EXPRESSION PROFILING FOR
IDENTIFYING MOLECULAR MARKERS USEFUL FOR
DIAGNOSIS OF METASTATIC CANCER**

5 CROSS REFERENCE TO RELATED PATENT APPLICATIONS

This application claims priority to U.S. Provisional Application Number 60/192,229 filed March 27, 2000, which is incorporated herein by reference.

FIELD OF THE INVENTION

10 The present invention relates to methods to isolate molecular markers that may be used to detect metastasized tumor cells.

BACKGROUND OF THE INVENTION

15 Throughout the specification references are cited in brackets. These references are incorporated by reference in their entirety to describe the state of the art.

20 Cancer represents a significant worldwide health problem. Cancer is an uncontrolled growth and spread of cells. For many cancers, metastasis to adjacent or distant tissues results in physiologic impairment and often death. Early diagnosis and the ability to diagnosis metastasis of primary tumors represent significant challenges in the effective treatment of neoplastic disease.

25 Stage at diagnosis is the single most important prognostic determinant for patients with cancer and dictates the role of adjuvant chemotherapy in this disease. Given the prognostic and therapeutic importance of staging, accurate histopathologic evaluation of lymph nodes to detect invasion by cancer cells is crucial. Specific diagnosis of cancer metastasis is currently preformed by histologic and cytologic

resemblance to normal tissue. Cancer cells frequently maintain their phenotypic characteristics of their normal cell of origin.

However, conventional microscopic lymph node examination has methodological limitations. Differentiation of single or even small clumps of tumor cells from other cell types can be difficult, limiting sensitivity. The standard practice of examining only several tissue sections from each lymph node can omit from review >99% of each specimen, introducing sampling error. These limitations are evident when the frequency of recurrence in patients with stage I and II colorectal cancer is considered. By definition, these patients do not have extra-intestinal disease at the time of curative resection. However, recurrence rates of 10% to 30% for lesions confined to the mucosa (stage I) and 30% to 50% for lesions confined to the bowel wall (stage II) have been reported.

Alternative methods to detect small numbers of tumor cells have been applied to staging, including intensive review of serial tissue sections, PCR to detect tumor-specific mutations, immunohistochemistry or and RT-PCR to detect the expression of biomarkers that are specifically expressed in cells that have undergone neoplastic transformation (Sloane, 1995, Lancet 345: 1255-6; Abati and Liotta, 1996, Cancer 78: 10-66). In some colorectal cancer studies, staging by these sensitive methods has correlated with disease. However, the labor- and cost-intensity of serial sectioning, the lack of uniform association between mutations and neoplastic transformation, and the lack of specificity of many biomarkers limit the applicability of these methods.

Easily detected molecular markers that are uniformly expressed by larger numbers of metastasized tumor would therefore be useful for metastasis detection and disease staging. Particularly needed is methodology to isolate useful molecule markers for the detection of metastatic tumor cells in tissues and/or bodily fluids. Such methodology would ideally be high throughput and utilize established robust protocols.

SUMMARY OF THE INVENTION

The present invention relates to a method for the isolation of tissue-specific molecular markers that are useful in the diagnosis of metastatic cancer.

One aspect of the invention is a method to identify molecular markers
5 useful for detecting tumor cells that have metastasized from an origin tissue to a destination tissue or fluid. The method comprises the steps of down-regulating in a population of origin tissue cells the activity of a transcription factor associated with terminal differentiation in the origin tissue, comparing an expression profile of the population of down-regulated origin cells with an expression profile of a population
10 of control origin cells, identifying candidate markers which are expressed in the population of control origin cells but not the population of down-regulated origin cells, and comparing expression of the candidate markers in populations of control origin cells, cancerous origin cells and destination cells, wherein a candidate marker
15 which is expressed in population of control origin cells and cancerous origin cells, but not the population of destination cells is a useful marker for the detection of cancer metastasized from the origin tissue to the destination tissue. The method may comprise the additional step of isolating the molecular marker. The method may also comprise the additional steps of identifying the transcription factor that binds to regulatory regions of a gene associated with terminal differentiation of the origin
20 tissue.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Functional characterization of deletion mutants of the human GC-C gene promoter. Deletion mutants of the GC-C gene 5'-flanking region
25 were linked to luciferase and co-transfected with the *Renilla* luciferase control plasmid pRL-TK into intestinal (T84, Caco2) and extra-intestinal (HepG2, HeLa, HS766T) cell lines. Data are expressed as luciferase activity relative to the pGL3 Basic promoterless construct (Relative Activity). Each bar represents the mean \pm the standard error of at least 3 independent transfections performed in duplicate.

Figure 2. DNase I protection of the proximal human GC-C promoter.

Footprinting reactions included the indicated mg quantities (NE) of HepG2 or T84 nuclear extract and the -46 to -257 promoter fragment labeled at the 5'-end of the coding strand. A control digestion contained 60 mg of bovine serum albumin (BSA).

- 5 Protected bases were identified by a Maxam-Gilbert sequencing reaction (G + A) of the labeled fragment. The sequence of FP1 is given. Arrowhead indicates DNase I hypersensitivity site at base -163.

Figure 3. Regulation of reporter gene expression by intestine-specific

- 10 protected elements. FP1 and FP3 were deleted from the -835 luciferase construct by *in vitro* mutagenesis, and wild-type and deletion constructs were expressed in HepG2 and T84 cells. Results are expressed as luciferase activity relative to a promoterless construct and represent the mean \pm the standard error of 3 independent transfections performed in duplicate.

Figure 4. Intestinal specificity of FP1 probe EMSA. Nuclear

- 15 extracts from intestinal or extra-intestinal cells, or BSA (10 mg), were incubated with labeled FP1 for 30 min. at room temp prior to separation on a non-denaturing 6% polyacrylamide gel.

Figure 5. Cdx2 binding element FP1 is required for GC-C reporter

- 20 gene activation. Putative binding sites for Cdx2 and HNF-4a are indicated on the -835 construct. T84 and HepG2 cells were transfected with the -835 reporter construct from which FP1 was deleted, or that construct containing the 'CCC' mutation. Results are expressed as (luciferase activity of mutant construct, luciferase activity of wildtype construct) x 100, and represent the mean \pm the standard error of 3 independent transfections performed in duplicate. The values expressed as relative
25 luciferase activities are, respectively, (wildtype; FP1 deletion; 'CCC' mutation): T84 (16.2 \pm 2.7; 1.9 \pm 0.3; 2.3 \pm 0.1) and HepG2 (2.1 \pm 0.1; 2.9 \pm 0.3; 2.2 \pm 0.1).

DESCRIPTION OF PREFERRED EMBODIMENTS

The present invention relates to methods to identify and characterize molecular markers useful for detecting metastasized tumor cells. Most commonly, molecule markers used to detect tumor cells are transcripts or proteins specifically expressed as a result of the hyperproliferative state of the cell. In contrast, the molecular markers that are identified and characterized by the method of the present invention are specifically expressed in terminally differentiated tissues and are not specific to tumor cells. Tumor cells continue to express the genes associated with terminal differentiation of their tissue of origin. The transcripts and proteins of these genes are ideally suited to detect tumor cells that have metastasized to a destination tissue, such as a lymph node, because the origin tissue specific markers will be out of place in the destination tissue. Because these molecular markers are specific to the origin tissue and not a particular tumor, they will broadly recognize many tumors metastasized from the origin tissue.

The method for identifying molecular markers useful for detecting metastasized tumor cells identifies "candidate" tissue-specific molecule markers and determines which of these candidate markers are suitable for the detection of metastatic cancer. Tissue-specific markers associated with the terminal differentiation of a desired origin tissue are characterized by down-regulating the activity of a transcription factor associated with terminal differentiation of origin tissue, comparing the expression profiles of the down-regulated origin tissue with unaltered control origin tissue, and identifying transcripts or proteins that are candidate tissue-specific markers by virtue of their expression being up- or down-regulated in conjunction with the down-regulation of the transcription factor. The expression of the candidate tissue-specific markers are compared in the control origin tissue, tumors derived from the origin tissue, and destination tissues of interest for biopsy. Candidate markers that are expressed in control origin tissue and tumors, but not destination tissue are useful markers for detecting metastatic tumor cells.

As used herein, the term "terminal differentiation" refers to a differentiation state of a cell or tissue from which no further differentiation can occur.

As used herein, the term “metastasize” refers to the process whereby cancer cells break loose from a tumor mass and form secondary tumors or metastases at other sites of the body.

As used herein, the term colorectal tract refers to the tissues and organs than comprise the large intestine to and including the rectum. These tissues and organs comprise, but are not limited to, the terminal ileum and ileocecal valve, the cecum, the ascending, transverse, descending and sigmoid colon, the rectum and the anal sphincters.

The origin tissue of the invention is any terminally differentiated tissue of the body in which tumor cells first arise. By “arise”, it is meant to confer to cells the hyperproliferative phenotype associated with tumor cells. The origin tissue is preferably a tissue from which cancer cells are most likely to metastasize. In a preferred embodiment, the tissue is mammalian, and in a most preferred embodiment, the tissue is human. In preferred embodiments, the origin tissue includes, but is not limited to, colorectal, intestine, stomach, liver, mouth, esophagus, throat, thyroid, skin, brain, kidney, pancreas, breast, cervix, ovary, uterus, testicle, prostate, bone, muscle, bladder and lung. It is particularly advantageous to use established cell lines in the method of the invention. The cell lines of particular interest represent terminally differentiated cells of the origin tissue, including embryonic tissue cell lines and immortalized cell lines (Yeager and Reddel, 1999, Curr. Opin. Biotechnology 10:465-469). Cell lines of particular interest include, but are not limited to, T84, Caco2, HT29, SW480, SW620, NCI H508, SW1116, SW1463, Hep G2, HS766T, and HeLa cells. These and additional cell lines of origin tissue may be obtained from the American Type Culture Collection (Manassas, VA), as well as from commercial sources.

Cancerous origin tissues are isolated from tumors that arise in the origin tissue. Cancerous cells may be obtained by removing tumors from patients. Established populations of tumor tissue, i.e. cell lines of tumor cells, can be used to advantage in the method of the invention. Cancer cell lines of interest include, but are not limited to, T84, Caco2, HT29, SW480, SW620, NCI H508, SW1116, SW1463,

Hep G2, HS766T, and HeLa cells. These cell lines and other useful cell lines may be obtained from the American Type Culture Collection (Manassas VA), as well as from commercial sources.

5 The destination tissue of the invention is any tissue or bodily fluid that may be biopsied to detect metastasized tumor cells. Several tissues of the body are well known to those in the art for their propensity to accumulate metastasized tumor cells, and these tissues are preferred for the destination tissue. However, the destination tissue may be any tissue of the body. Destination tissues of particular interest include, but are not limited to, lymph node, blood, cerebral spinal fluid, and
10 bone marrow. Additional cell lines for origin tissue cells may be obtained from the American Type Culture Collection (Manassas, VA), as well as from commercial sources. Preferably, biopsy or resected tissue is used as the destination tissue.

 The transcription factors used in the method of the invention are transcription factors that are associated with terminal differentiation of the origin
15 tissue. Many such transcription factors are already known to those skilled in the art. In preferred embodiments, the transcription factor is associated with the terminal differentiation of a preferred origin tissue. In preferred embodiments, the transcription factors include, but are not limited to, Cdx2 (intestine) (Mallo, G.V. et al., 1997 Int J Cancer 74:35-44; Genbank Accession No. BF591065), STAT5 (breast)
20 (Hou, J. et al., 1995 Immunity 2:321-329; Genbank Accession No. L41142), NKX3.1 (prostate) (Genbank Accession No. AF247704), GBX2 (prostate) (Lin, X. et al., 1996 Genomics 31: 35-342; Genbank Accession No. NM U13219), FREAC-2 (lung) (Pierrou, S. et al., 1994 EMBO J. 13:5002-5012; Genbank Accession No. U13220), Pit1 (thyroid) (Wu, W. et al., 1998 Nat Genet 18:147-9; Genbank Accession No. NM
25 006261) HNF4 (liver) (Chartier, F.L. et al., 1994 Gene 147:269-272; Kritis, A.A. et al., 1996 Gene 173:275-80; Genbank Accession Nos. X76930, X87870, X87872, X87871), LFB1 (liver) (Bach, I. et al., 1990 Genomics 8:155-164; Genbank Accession No. NM 000545), IPF1 (pancreas) (Stoffel, M. et al., 1995 Genomics 28:125-126; Genbank Accession Nos. NM 000209, U30329), Isl1 (pancreas) (Wang,
30 M. and Drucker, D.J., 1994 Endocrinology 134:1416-1422; Genbank Accession Nos.

XM 003669, NM 002202) and MyoD (muscle) (Pearson-White, S.H., 1991 Nucleic Acids Res. 19:1148; Genbank Accession No. X56677), all of which are incorporated by reference herein.

- The method of the present invention may, in some embodiments,
- 5 further comprise steps to identify a transcription factor gene associated with terminal differentiation. These additional steps comprise identifying the transcription factor that binds to the regulatory regions of a gene associated with terminal differentiation in the origin tissue. There are many protocols currently available and known to those skilled in the art to characterized transcription factors and transcription factor genes.
- 10 In a preferred embodiment, electromobility shift assays and/ or supershift assays are used to characterize the transcription factor that binds to the regulatory region of a gene whose expression is associated with terminal differentiation. Example 1 illustrates the characterization of transcription factor Cdx2 by its binding to the regulatory regions of the gene encoding the intestine-specific protein guanylyl
- 15 cyclase C.

- In the method of the invention, the activity of transcription factor associated with terminal differentiation is "down-regulated" in a population of origin tissue cells. By "down-regulated", it is meant that the activity of the transcription factor is reduced in the cell population as compared to a "normal" or control cell
- 20 population. As used herein, a "cell population" refers to a cell culture, tissue culture, resected tissue or biopsy sample, or any group of cells from the desired tissue type. A population of normal or control origin cells refers is a population of origin cells from the culture of origin tissue cells used for down-regulating the transcription factor, but without modification of the activity of the transcription factor.

- 25 The activity of the transcription factor may be down-regulated in cell populations by several means well known to those in the art. In some embodiments, the transcription factor gene is down regulated by site-directed mutagenesis of the coding or regulatory regions of the gene, or the transcription of an antisense gene constructed from the coding sequence of the transcription factor gene. Alternately, in
- 30 other embodiments, the activity of the transcription factor is blocked or inhibited by

specific antibodies, DNA-binding molecules, or small molecules that interfere with the activity of the transcription factor by interfering with the assembly and/or initiation of the transcriptional complex. Inhibitor polynucleotide molecules of interest include, but are not limited to, FP1, FP1B and SIF1 (see Example 1). Finally,

- 5 in other embodiments, the transcription factor may be down-regulated by activating a signaling event that inactivates the transcription factor, such as the addition of an extracellular ligand that initiates a cell-signaling event that phosphorylates and inactivates the transcription factor. These methods will be well known by those skilled in the art, and protocol can be found in many laboratory manuals, such as
- 10 Ausubel et al. **Current Protocols in Molecular Biology**. New York: John Wiley & Sons, Inc., 2000. These embodiments are meant to illustrate methods by which to generate down-regulated origin cells. Other manners of down-regulation will be well known to those skilled in the art and are included in the scope of the method of the present invention.

- 15 In a preferred embodiment, the down-regulated origin cells are *cdx2*-null polyps. *Cdx2*-null polyps can be resected from a mouse that is heterozygous for an inactive copy of the homeobox gene *cdx2*, which controls cell differentiation in the intestinal epithelium (Chawengsaksophak et al., 1997, Nature 386:84-87; Tamai et al., 1999, Cancer Res. 59:2965-2970; Beck et al., 1999, PNAS 96:7318-7323;
- 20 incorporated by reference herein). *Cdx2* stimulates the markers of endocyte differentiation. These heterozygous mice develop multiple intestinal polyp-like lesions that do not express active *Cdx2* and the *Cdx2*-related markers. In this embodiment, the comparison of the expression profiles of *Cdx2*-null polyps with surrounding intestinal tissue will identify the *Cdx2* stimulated markers of endocyte
- 25 differentiation.

- The method of the invention comprises the step of comparing the expression profile of the population of down-regulated origin cells with the expression profile of the population of control origin cells. By "expression profile" it is meant the array of nucleic acids or proteins that are expressed in a cell population. Most
- 30 commonly, expression profiles are arrays of nucleic acid molecules, primarily mRNA

- molecules, that are found in the profiled cell population. Methods to compare RNA expression profiles are well known to those in the art. Some methods of particular interest include, but are not limited to, differential display (Welsh et al., 1992, *Nucleic Acids Res.* 20:4695-4970; Liang and Pardee, 1992, *Science* 257:967-970; Barnes, 5 1994, *Proc. Natl. Acad. Sci. USA* 91:2216-2220; Cheng et al., 1994, *Proc. Natl. Acad. Sci. USA* 91: 5695-5699; and the references cited therein), subtractive hybridization (Diatchenko et al., 1996, *Proc. Natl. Acad. Sci. USA* 93:6025-6030; Gurskaya et al., 1996, *Anal. Biochem.* 240:90-97; Endege et al., 1999, *Biotechniques* 26: 542-550; and the references cited therein), expression arrays (Schena et al., 1995, 10 *Science* 270: 467-470; Shalon et al., 1996, *Genome Res.* 6: 639-645; Cheung et al., 1999, *Nature Genetics* 21(Suppl.): 15-19; and the references cited therein), Serial Analysis of Gene Expression (SAGE) (Velculescu et al., 1995, *Science* 270: 484-487; Zhang et al., 1997, *Science* 276: 1268-1272; Adams et al., 1996, *Bioessays* 18: 261-262; and the references cited therein), Rapid Analysis of Gene Expression (RAGE) 15 (Wang et al., 1999, *Nucleic Acids Res.* 27: 4609-4618; and the references cited therein), Massively Parallel Signature Sequencing (MPSS) (Brenner et al., 2000, *Nature Biotech.* 18: 630-634; and references therein) and Tandem Arrayed Ligation of Expressed Sequence Tags (TALEST) (Spinella et al., 1999, *Nucleic Acids Res.* 27: e22 (I-VIII); and references therein).
- 20 Many of the aforementioned techniques may be preformed using commercially available kits, reagents and apparatuses. Commercial kits for differential display may be purchased, such as the Delta® Differential Display Kit (Clontech, Palo Alto, CA), among others. Commercial kits for subtractive hybridization may be purchased, such as Clontech PCR-Select® Subtraction 25 (Clontech, Palo Alto, CA), among others. Micro-arrays of popular cDNA populations may be purchased (Incyte Genomics, Inc, St. Louis. MO), or custom micro-arrays may be ordered from commercial sources (Radius Biosciences, Medfield MA ; ProtoGene Laboratories, Inc., Menlo Park CA). A preferred membrane-format microarray is LifeGrid™ Sequence-Verified Gene Expression Array Kits (Incyte 30 Pharmaceuticals, Inc., St. Louis, MO) and a preferred slide-format microarray is

GEM® Gene Expression Microarray (Incyte Pharmaceuticals, Inc., St. Louis, MO). Commercial kits for RAGE are available from Kirkegaard & Perry Laboratories, Inc. (Gaithersburg, MD). GeneTag®, a proprietary technology developed by Celera Genomics (Rockville, MD), may also be used to quantify gene expression in a profile of RNA transcripts.

Protein expression profiles may also be compared by methods that will be well known to those in the art. Methods of particular interest include, but are not limited to, 2-Dimensional Electrophoresis - Mass Spectroscopy (2DE-MS) (O'Farrell, 1975, J. Biol. Chem. 250: 4007-4021; Patterson and Aebersold, 1995, Electrophoresis 16: 1791-1814; Gygi et al., (2000) Curr. Opinion in Biotech. 11: 396-401; and references cited therein) and Isotope-Coded Affinity Tags (ICAT) (Gygi et al., 1999, Nature Biotech. 17: 994-999; Gygi et al., 2000, Curr. Opinion in Biotech. 11: 396-401; and references cited therein).

Nucleic acid molecules or protein molecules of interest identified by the comparison of expression profiles may additionally be isolated using methods that will be well known to those skilled in the art. The isolation method chosen depends in many cases on the method used to compare the expression profiles, and the preferred method will often be described in the reference that describes the method of comparison (see aforementioned citations). For example, nucleic acid bands may be removed from a polyacrylamide gel, agarose gel or nitrocellulose, the nucleic acids eluted and cloned using techniques well known in the art (Ausubel et al. **Current Protocols in Molecular Biology**. New York: John Wiley & Sons, Inc., 2000).

The method of the invention comprises the step of comparing the expression of the candidate markers in several kinds of cells. There are many methods to compare the expression of single genes which will be well know to those in the art (Ausubel et al. **Current Protocols in Molecular Biology**. New York: John Wiley & Sons, Inc., 2000), including but not limited to, northern analysis, Southern analysis with cDNA, RNase protection assays, quantitative PCR, competitive PCR, 5' nuclease assays (Lie and Petropoulos, 1998, Curr. Opin. Biotech. 9:43-48 and the

references cited therein), western analysis, dot blot western, ELISA and other immunoassays, and immunohistochemistry.

The molecular markers identified by the method of the invention may be used to diagnose and stage cancer in mammalian patients, including following the development of recurrence of cancer after surgery and screening normal patients for the development of cancer. In the case of cancer patients, the molecular markers utilized would be identified ideally from the same tissue that the patients cancer arose. In the case of patients without a history of cancer, a selection of molecular markers isolated from different origin tissues is preferred. The metastases may be diagnosed by any technique that will detect the nucleic acid or protein molecular marker. The sensitivity of the technique will determine in part the size of metastasis that can be detected. Preferred techniques utilize PCR, ELISA, and the like. Example 2 illustrates a particularly preferred method to diagnose metastasized cancer with the molecular markers of the method.

Tissue specific molecular markers can also be utilized to localize therapeutics to specific tissue and organ systems. This use is particularly appropriate for tissue-specific molecular markers that are localized on the surface of the tissue cells. These therapeutics include, but are not limited to, chemotherapeutics, analgesics, antibiotics, anti-inflammatories, hormones and stimulants.

Protein molecular markers may be used to generate antibodies that may be used in diagnosis method and to localize therapeutics. Polyclonal antibodies and monoclonal antibodies, and fragments thereof, and various conjugates of them can be made by methods well known in the art.

Example 1 Cdx2 is a Transcription Factor Associated with the Intestinal-Specific Expression of Guanylyl Cyclase C

This illustrates the identification of a transcriptional activating factor required for intestine-specific expression of guanylyl cyclase C (GC-C). A region of the proximal GC-C promoter required for specific expression in intestinal cells that contains a protected region, FP1, with a consensus binding sequence for Cdx2. FP1

formed a complex specifically with nuclear proteins only from intestinal cells, and this complex was recognized by anti-Cdx2 antibody. Elimination or mutation of the Cdx2 consensus binding sequence within FP1 reduced reporter gene activity in intestinal cells to that obtained in extra-intestinal cells. These data suggest that Cdx2
5 activates tissue-specific transcription of GC-C.

Materials and Methods

Genomic Library Screening and Sequencing. The GC-C gene 5' regulatory region was cloned from a λ FIXII human genomic library (Stratagene, La Jolla CA). The library was screened by hybridization with a probe specific for exon 1
10 of the guanylyl cyclase C (GC-C) cDNA. A 2.8 kb XbaI fragment that included 2 kb upstream of the start site of transcription was subcloned into Bluescript KS (Stratagene). All constructs were generated from this Bluescript/human GC-C gene construct. The nucleic acid sequence of each construct was confirmed by BigDye terminator[®] reaction chemistry for sequence analysis on the Applied Biosystems
15 Model 377 DNA sequencing systems (Perkin-Elmer, Norwalk CN; Applied Biosystems, Foster City CA).

Reporter Gene Constructs. Fragments -835 to +117, -257 to +117, -129 to +117, and -46 to +117, relative to the start site of transcription, were isolated from Bluescript KS constructs by digestion with selected restriction endonucleases
20 (Mann et al., 1996, Biochim Biophys Acta 1305:7-10). These fragments were blunt-ended and ligated into the EcoRV site of Bluescript KS. Inserts were excised from Bluescript KS with SmaI and KpnI and ligated into the pGL3-Basic Luciferase Vector (Promega, Madison WI). The pGL3 Control Vector containing an SV40 promoter with enhancers, was used as a positive control.

25 Mutations were created in the -835 to +117 pGL3 construct utilizing the PCR-based Ex-site Mutation Kit (Stratagene). Deletion constructs were created using primers flanking the sites of interest. The FP1 "CCC" mutant was created using the phosphorylated primers:

30 5' GCCCATAGCTCTGACCTTTCTG 3' (SEQ ID NO:1) and
5' AGAGAGATTAGCTGGGCCTCACCC 3' (SEQ ID NO:2).

Cell Culture and Transfection. All cell lines were obtained from American Type Culture Collection (Rockville, MD). T84 cells were grown in DMEM/F12 (Life Technologies, Rockville MD), Caco2 cells in DMEM (Life Technologies), HepG2 and HS766T cells in DMEM High Glucose (Cellgro®, Mediatech, Inc., Herndon VA), and HeLa cells in MEM with glutamine (Life Technologies). All cell lines were maintained at 37°C in a 5% CO₂/95% air atmosphere and passaged every four days. Assays of reporter gene activity were conducted with cells plated in 6-well seeded at either 5.0×10^5 (T84, Caco2, and HeLa) or 1.0×10^6 cells per well (HepG2 and HS766T). Cells were incubated overnight, washed one time with PBS, and supplemented with fresh media before transfection.

Plasmids purified with the Qiafilter Kit (Qiagen, Valencia CA) were transfected into cells with the non-liposomal lipid transfection reagent Effectene® (Qiagen). All cell lines were co-transfected with both 0.4 mg of firefly luciferase experimental reporter constructs, modified from pGL3-Basic, and 0.1 mg of the *Renilla* luciferase control reporter, pRL-TK, driven by a viral thymidine kinase promoter (Promega). Cells were incubated with transfection complexes for 24 h, rinsed with PBS, then supplemented with appropriate media and incubated for a further 24 h. After a total of 48 h, cells were lysed and assayed using the protocol and materials in the Dual-Luciferase Reporter Assay system (Promega). Luminescence was measured with a BioOrbit 1251 Luminometer (Pharmacia LKB, Uppsala Sweden). Luciferase expression from pGL3 constructs was normalized to pRL-TK expression.

Nuclear Protein Extraction. Nuclear extracts were prepared essentially as previously described (Ausubel et al. Current Protocols in Molecular Biology. New York: John Wiley & Sons, Inc., 2000). Nuclear protein concentration was determined using Coomassie Protein Assay Reagent (Pierce, Rockford IL).

DNase I Footprinting. A fragment of the GC-C gene regulatory region -46 to -257 relative to the start of transcription was obtained by digestion with DraIII and AflII, blunt-ended, and subcloned into the Bluescript® KS EcoRV site, as

described above, and then digested with EcoRI and HindIII to ensure that the coding strand of the probe was singly end-labeled with [α - 32 P]dCTP. Products obtained from footprinting reactions were separated on a denaturing 6% polyacrylamide gel and visualized by a Phosphorimager SI (Molecular Dynamics, Sunnyvale, CA).

- 5 **Electromobility Shift Assay (EMSA).** Protein-DNA binding reactions performed in the same buffer as the DNase I protection assay (4% glycerol, 10 mM Tris-HCl (pH 7.5) 50 mM NaCl, 2.5 mM MgCl₂ and 5 mM DTT) included 1 mg of Poly(dI-dC)-Poly(dI-dC) (Amersham Pharmacia Biotech, Piscataway, NJ) and 30 kcpm of probe. Reactions were initiated by the addition of nuclear extract and
- 10 incubated for 30 min at room temp to produce protein complexes which were separated on a 6% non-denaturing, polyacrylamide (37.5:1) gel in 0.5 x TBE running buffer. Gels were dried prior to visualization of radiolabelled complexes by autoradiography. In competition assays, unlabelled competitor was added to the reaction mixtures at concentrations ranging from 25-fold to 250-fold molar excess of
- 15 the labeled probe prior to the addition of the nuclear extract. Supershift assays were performed by adding 2 ml of murine Cdx2 antibody after an initial incubation period of 30 min; incubation was then continued for an additional 30 min. Transcribed and translated murine Cdx2 protein was generated *in vitro* using linearized pRc/CMV-Cdx2 expression vector as a template for the TNT-Quickcoupled Kit
- 20 (Promega).

Oligonucleotide probes for EMSA were synthesized. Complementary oligonucleotides in 10 mM Tris-HCl (pH 7.5), 1mM EDTA were annealed in a Hybaid Thermal Cycler by a programmed ramp in temp from 95°C to 25°C over the course of 1 h. The single stranded sequences of the probes were:

- 25 FP1: 5' CAGCTAATCTCTCTGTTTATAGCTCTGACCTTTC 3' (SEQ ID NO:3)
- FP1B: 5' ATCTCTCTGTTTATAGCTCTGACCTTCTGGGTGC 3' (SEQ ID NO:4)
- FP1-CCC: 5' CAGCTAATCTCTCTGCCCATAGCTCTGACCTTTC 3' (SEQ ID NO:5)
- 30

SIF1: 5' GATCCGGCTGGTGAGGGTGCAAT**AAAACTTTAT**GAGTA 3'
(SEQ ID NO:6)

Bolded sequences indicate specific Cdx2 binding sites. A mutation created in the FP1 protected site is underlined. Five pmol of annealed oligonucleotide probe were end-labeled employing 1 unit of T4 polynucleotide kinase and 2 ml of 7,000 Ci/mmol [γ - 32 P]ATP (Ausubel et al. Current Protocols in Molecular Biology. New York: John Wiley & Sons, Inc., 1999). Labeled probes were purified over Qiaquick nucleotide purification columns (Qiagen).

Southwestern and Western Blotting. Nuclear extracts were denatured in reducing SDS sample buffer, separated on an 8% Tris-glycine-SDS polyacrylamide gel, and transferred to nitrocellulose. For Southwestern analysis, the blotted proteins were blocked for 1 h at 4° in Z' buffer (25 mM Hepes-KOH (pH 7.6), 12.5 mM MgCl₂, 20 % glycerol, 0.1% Nonidet P-40, 100 mM KCl, 10 mM ZnSO₄, 1 mM DTT) containing 3% non-fat dry milk (Hames and Higgins. Gene Transcription: A Practical Approach. The Practical Approach Series. New York: Oxford University Press, 1993.). The membrane was rinsed for 5 min in EMSA binding buffer and hybridized with 20 ml of EMSA binding buffer with 100 kcpm/ml of labeled FP1 probe for 1 h at room temp. The membrane was then washed for 5 min each in three changes of EMSA binding buffer, dried and visualized by autoradiography.

Western blots were blocked in TBS/0.1% Tween-20 with 5% non-fat dry milk, and probed with Cdx2 antibody diluted 1:5000. Binding of primary antibody was visualized using goat anti-rabbit alkaline phosphatase-conjugated secondary antibody diluted 1:10,000 (Sigma). Alkaline phosphatase substrates BCIP and NBT were used in an AP Color Kit (Biorad).

Results

Determination of elements controlling intestine-specific expression in the 5' regulatory region of the GC-C gene. Minimal luciferase activity was obtained when various cell lines were transfected with the -46 construct (Fig. 1). In contrast, luciferase activity increased in intestinal cells transfected with each of the other reporter gene constructs (Fig. 1). Luciferase activity did not increase when

extra-intestinal cells were transfected with these constructs (Fig. 1). These results are consistent with previous studies of GC-C gene regulation, and suggest that there are one or more tissue-specific regulatory elements within the +118 to -257 region. 12 Since transfection with the -46 to -129 construct resulted in a significant increase in activity of the reporter gene in intestinal cells only (Fig. 1), and since this region is highly conserved evolutionarily, it was chosen for detailed structure-function analysis.

DNAse I protection by intestine-specific nuclear protein binding to the 5' regulatory region of GC-C. DNAse I protection assay revealed two regions (-75 to -83, FP1; -164 to -178, FP3) which were protected only by nuclear extracts from intestinal cells (T84; Fig. 2). Regions -104 to -137 (FP2) and -180 to -217 (FP4) were protected by nuclear extracts from either intestinal (T84) or extra-intestinal (HepG2) cells, although the proximal and distal ends of FP2 exhibited different patterns of protection. These data suggest that the protected regions designated FP1 and FP3 were specific binding sites for nuclear proteins from intestinal cells. In addition, an intestine-specific site of open chromatin structure in the proximal 5'-flanking region of the GC-C gene was identified by a DNAse I hypersensitive site at base -163 (Fig. 2).

Transcriptional activity of the -857 construct following deletion of FP1 or FP3. Transfection of T84 cells revealed that deletion of FP3 increased luciferase activity 2.5-fold relative to the wild-type construct (Fig. 3). In contrast, elimination of FP1 reduced luciferase activity in T84 cells to levels observed in HepG2 cells (Fig. 3). These data suggest that FP3 contains a negative regulatory element, and that FP1 contains an intestine-specific positive regulatory element. Analysis by TRANSFAC (Heinemeyer et al., 1998, Nucleic Acids Res. 26: 364-370), a database of transcription factor binding sites, revealed that FP1 contains the consensus binding site for the homeodomain protein Cdx2 (Quandt et al., Nucleic Acids Res 1995; 23:4878-84). Since Cdx2 is a transcription factor that directs intestine-specific expression of several genes, FP1 was more closely examined (Traber and Silberg, 1996, Annu Rev Physiol 58:275-97).

Specific complexes are formed by intestinal nuclear extract and

FP1 probe. The ability of the protected site FP1 to form intestine-specific complexes was determined by incubating an oligonucleotide probe with nuclear extracts prepared from T84, Caco2, HepG2, or HeLa cells. Indeed, several complexes were obtained by

- 5 EMSA when the FP1 probe was incubated with nuclear extracts from those cells (Fig. 4). However, only one complex satisfied criteria for intestinal specificity, including formation by nuclear extracts from T84 and Caco2 cells, but not from HepG2 or HeLa cells. Extracts from T84 and Caco2 cells, but not from HepG2 or HeLa cells, also formed complexes with SIF1 that were identical to those obtained previously with that
- 10 probe, demonstrating the integrity of the extracts (Suh et al., 1994, Mol Cell Biol 14:7340-51). All of the EMSA complexes formed with T84 nuclear extracts were competed with increasing amounts of unlabelled FP1 probe in a concentration-dependent manner. In contrast, an unlabelled competitor in which the Cdx2 binding site was specifically mutated (FP1-CCC probe, see Materials and
- 15 Methods) did not compete against the intestine-specific complex. SIF1, an oligonucleotide containing two consensus binding sites for Cdx2, selectively prevented the formation of the FP1-dependent intestine-specific complex with greater potency than unlabelled FP1, but generally did not affect the binding of the remaining T84-EMSA complexes (Suh et al., 1994). These data suggest that the
- 20 intestine-specific factor that binds to the FP1 protected site is most likely Cdx2.

- Cdx2 binds specifically to the FP1 probe.** To determine whether FP1 is a binding site for Cdx2, labeled FP1 was incubated with *in vitro* transcribed and translated murine Cdx2. This resulted in a complex whose mobility was identical to the intestine-specific complex formed by T84 nuclear extract. In contrast, labeled
- 25 FP1-CCC did not form the intestine-specific complex with either Cdx2 or T84 nuclear extract. An antibody against Cdx2 decreased the mobility of the specific complex formed between labeled FP1 and either T84 nuclear extract or *in vitro* transcribed and translated Cdx2. In contrast, an antibody against a related homeodomain transcription factor, Cdx1, did not alter the mobility of the intestine-specific complex. These data
- 30 lead to the conclusion that the FP1 protected site is a binding site for Cdx2.

Identification of the intestine-specific nuclear factor by

Southwestern and Western blots. Whether the FP1 probe and anti-Cdx2 antibody bound to the same intestine-specific protein was examined. Labeled FP1B, which is highly homologous to FP1 probe, specifically bound to an intestine-specific protein of ~40 kDa in T84 and Caco2, but not HepG2, nuclear extracts. In addition, FP1B probe bound to a ~131 kDa protein present in all cell lines examined. Similarly, anti-Cdx2 antibody recognized a protein doublet of ~40 kDa expressed in T84, but not in HepG2 or HeLa, cell nuclear extracts, a pattern which is characteristic of Cdx2 (James et al., 1994, J Biol Chem 269:15229-37). Thus, the FP1 protected region binds to an intestine-specific factor of the same molecular weight and antigenic recognition as Cdx2. Furthermore, Southwestern blots revealed that FP1 probe binds directly to Cdx2.

Role of the Cdx2 binding element (FP1) in intestine-specific gene expression of the GC-C promoter. The 'CCC' mutation was introduced into the FP1 element of the -835 luciferase reporter gene construct. This mutated reporter gene construct exhibited reduced activity in T84 cells that was comparable to the construct from which the entire FP1 region was deleted (Fig. 5). Neither the FP1 deletion nor the 'CCC' mutation in FP1 altered luciferase expression in HepG2 cells (Fig. 5). These data demonstrate that an intact Cdx2 binding site is required for activity of the GC-C promoter. Indeed, disruption of the Cdx2 binding site resulted in minimal activity.

Example 2 Guanylyl Cyclase C Messenger RNA is used as a Molecular Marker to Detect Recurrent State II Colorectal Cancer

This example illustrates the use of a tissue-specific molecule marker to diagnose metastases. Detection of GCC mRNA by RT-PCR enhances the accuracy of colorectal cancer staging. The expression in lymph nodes of GCC mRNA, a molecular marker for colorectal cancer cells in extraintestinal tissues, is associated with disease recurrence in patients with histologically negative nodes (stage II).

Expression of GCC mRNA reflects the presence of colorectal cancer micrometastases below the limit of detection by standard histopathology. GCC-specific RT-PCR can reliably and reproducibly detect a single human colorectal cancer cell (T84 cells, ATCCC, Rockville, MD) in 10^7 nucleated blood cells (Carrithers et al., 1996, Proc Natl Acad Sci USA, 93:14827-32).

GCC, a member of the guanylyl cyclase family of receptors, is specifically expressed only in intestinal mucosal cells. However, GCC expression persists in intestinal cells that undergo neoplastic transformation to colorectal cancer cells. Examination of >300 surgical specimens demonstrated that GCC was specifically expressed by all primary and metastatic colorectal cancer cells, but not by any other extraintestinal tissues or tumors. GCC is identified only in lymph nodes from stage II patients who suffered recurrence ≤ 3 y, but not in lymph nodes from patients without recurrent disease ≥ 6 y, following diagnosis.

Materials and Methods

Patients and tissues. The Thomas Jefferson University Hospital tumor registry database was examined for patients who had undergone treatment for colorectal cancer between 1989 and 1995, an interval permitting adequate follow-up of patients for this study. This initial search was designed to exclude patients with recurrent disease >3 y following index surgery to avoid inadvertent inclusion of patients with metachronous, rather than recurrent, cancer. This search yielded 445 patients with invasive colon or rectal carcinoma with no evidence of metastases (N_0M_0) at the time of surgery. Of these, 260 patients underwent surgery at Thomas Jefferson University that yielded lymph nodes. Subsequently, 167 patients were excluded because they had TNM stage I disease or less (T_0 , T_1 or $T_2N_0M_0$), developed recurrent disease locally or at unspecified sites, or received neoadjuvant chemo- or radiotherapy. Fifty-six patients with no evidence of recurrence were then excluded because they had <6 y of follow up. After these exclusions, a total of 18 patients with no evidence of disease for ≥ 6 y following surgery and considered clinically cured remained. These patients formed the control group. Similarly, all 19 patients who developed metastases ≤ 3 y following surgery were included in the case group.

Sixteen patients in the control group and 12 patients in the case group had pathology specimens available for further analysis. Two patients in the control group (patients 9 and 16; 12.5%) and 1 patient in the case group (patient 24; 8.3%) received 5-fluorouracil-based adjuvant chemotherapy following surgery.

- 5 **Reverse transcriptase-polymerase chain reaction.** Preliminary studies demonstrated that mRNA isolated from 10 μ m sections from individual lymph nodes yielded insufficient RNA for RT-PCR analyses. Consequently, at least five 10 μ m sections of representative lymph nodes for each patient were pooled and deparaffinized, and the total RNA isolated (Waldman et al. 1996, Dis Colon Rectum 41:1-6.). RT-PCR was performed employing RNA PCR kit ver.2 (Takara Shuzo Co., Ltd., Kyoto, Japan; Carrithers et al., 1996, Proc Natl Acad Sci USA 93:14827-32; Waldman et al., 1996, Dis Colon Rectum 41:1-6). Only total RNA that yielded amplicons following β -actin-specific RT-PCR was employed in studies outlined below. GCC-specific and nested carcinoembryonic antigen-specific RT-PCR was
- 10 performed as described previously (Carrithers et al., 1996, Proc Natl Acad Sci USA 93:14827-32; Waldman et al., 1996, Dis Colon Rectum 41:1-6; Liefers et al., 1998, New Engl J Med 1998;339:223-8). RT-PCR reactions were separated by electrophoresis on 4 % NuSieve 3:1 agarose® (FMC Bioproducts, Rockland, ME) and amplification products visualized by ethidium bromide. Positive controls, consisting
- 15 of RNA isolated from human colorectal cancer cells expressing GCC and carcinoembryonic antigen (Caco2 cells; American Type Culture Collection, Rockville, MD) and negative controls, consisting of incubations in which no template was added and RNA from lymph nodes devoid of colorectal cancer, were included. Amplicon identity was confirmed by sequencing. Production of GCC-specific
- 20 amplicons was confirmed by Southern analysis, employing a 32 P-labeled antisense probe complimentary to a sequence internal to primers used for amplification (Kroczeck, 1993, J Chromatog 618:133-145).
- 25

Statistical analysis. Results are expressed as the mean \pm SD except disease-free and overall survival, which are expressed as the median \pm range. P

values were calculated using Fisher's Exact test. The odds ratio with exact 95% confidence interval (CI) was calculated employing the StatXact 4.0 statistical software package (CYTEL Software Corp., Cambridge, MA).

Results

5 **Characteristics of patients evaluated by RT-PCR.** The age of patients ranged from 37 to 85 y (68.1 ± 9.5 y). The ages of females (range = 52 - 85 y; 64.5 ± 10.5 y) and males (range = 37 - 82 y; 70.9 ± 7.8 y) were similar. The ratio of males to females was balanced between control (8:9) and case (5:7) groups. One female patient was African-American; all other patients were Caucasian. The ratio of

10 T₃ to T₄ disease was 3:13 in the control group and 4:8 in the case group. Patients were followed for 9 to 105 months (67.4 ± 30.7 months). Patients in the control group were followed for 73 to 105 months (89.9 ± 7.8 months) while those in the case group were followed for 9 to 78 months (37.3 ± 22.6 months). In the control group, one patient (6.3%) developed a new primary colonic lesion 96 months after initial

15 diagnosis, one (6.3%) died of causes unrelated to colorectal cancer, and the remaining 14 (87.5%) were alive and free of disease 88 (range, 73-97) months following diagnosis. In the case group, 8 (66.6%) patients died of recurrent colorectal cancer following intervals of disease-free and overall survival of 13 (range, 3-35) and 19 (range, 9-64) months, respectively. Four (33%) were alive with metastases following

20 intervals of disease-free and overall survival of 12 (range, 2-36) and 52 (range, 17-78) months, respectively.

RT-PCR analysis of RNA expression in lymph nodes. For the 28 patients in the control and case groups, a total of 524 (18.4 ± 12.5 lymph nodes/patient) lymph nodes collected at surgery were reported free of tumor by

25 histologic review. The number of lymph nodes obtained from each patient at the time of initial operative staging was similar between control (19.9 ± 13.2) and case (17.2 ± 12.7) groups. Twenty-one patients (75%) yielded 159 paraffin-embedded lymph nodes (7.6 ± 5.2 lymph nodes/patient) that could be adequately evaluated by RT-PCR. Lymph nodes omitted from RT-PCR analysis were not available from pathology (326

30 lymph nodes from 28 patients; 62.2% of 524 lymph nodes obtained at surgery) or did

not yield RNA (39 lymph nodes from 7 patients; 7.4% of 524 lymph nodes obtained at surgery; 19.7% of 198 lymph nodes available for RT-PCR analysis). The number of lymph nodes available for RT-PCR analysis was balanced between control (6.4 ± 3.0) and case (8.1 ± 6.3) groups.

- 5 β -Actin-specific amplicons (an indicator of intact RNA) were not detected in total RNA from pooled sections of lymph nodes of 5 (41.7%) patients from the case group and 2 (16.7%) patients from the control group and these patients were excluded from further analysis. Total RNA extracted from pooled lymph node sections from the remaining 21 patients was analyzed by RT-PCR using GCC-specific
- 10 primers. GCC-specific amplicons were not detected in any reaction using RNA from lymph nodes of patients in the control group ($p=0.004$; Table 1). The absence of GCC-specific amplicons in these reactions was confirmed by Southern analysis and suggests the absence of colorectal cancer micrometastases in lymph nodes of patients free of disease. In contrast, GCC-specific amplicons were detected in all reactions
- 15 using RNA from lymph nodes of patients in the case group (Table 1). The presence of GCC-specific amplicons in these reactions was confirmed by sequencing and/or Southern analyses and suggests the presence of colorectal cancer micrometastases in lymph nodes of patients with recurrent disease. Of note, GCC mRNA was not expressed in any of 39 lymph nodes from 21 other patients without colorectal cancer
- 20 (negative controls) that have been analyzed by RT-PCR to date.

Table 1. GCC mRNA expression in lymph nodes and patient outcome.

	Patient	GCC mRNA ^a	DFI ^b	OS ^c	Vital Status
	Controls				
5	6	(-)	97	97	Alive, NED ^d
	7	(-)	96	105	Alive, New 1 ^o Colon Cancer (T ₃ N ₁ M ₁)
	8	(-)	96	96	Alive
	9	(-)	82	82	Alive
	10	(-)	86	86	Died of Dehydration
10	11	(-)	89	89	Alive
	12	(-)	94	94	Alive
	13	(-)	87	87	Alive
	14	(-)	86	86	Alive
	15	(-)	87	87	Alive
15	16	(-)	73	73	Alive
	Cases				
	17	(+)	13	15	Dead 2 ^o to Liver Metastases
	18	(+)	15	52	Dead 2 ^o to Liver Metastases
	19	(+)	3	9	Dead 2 ^o to Liver Metastases
20	20	(+)	14	20	Dead 2 ^o to Liver Metastases
	21	(+)	2	78	Alive with Liver Metastases
	22	(+)	12	25	Alive with Liver Metastases
	23	(+)	9	55	Dead 2 ^o to Lung and CNS Metastases
	24	(+)	29	64	Alive with Lung and Bone Metastases
25	25	(+)	17	19	Dead 2 ^o to Liver, Lung and Bone Metastases
	26	(+)	11	17	Alive with Lung Metastases

^aGCC mRNA detected (+) or absent (-) in lymph nodes.^bDisease-free interval (months after diagnosis).^cOverall Survival (months after diagnosis).^dNED, no evidence of disease.

30

Carcinoembryonic antigen is a glycoprotein expressed by <60% of colorectal cancers and by other tumors, normal cells, and in some non-malignant pathological conditions. RT-PCR analysis of carcinoembryonic antigen expression has been suggested to be a marker of colorectal cancer micrometastases in lymph nodes. In the present study, total RNA extracted from pooled lymph node sections was analyzed by RT-PCR using carcinoembryonic antigen-specific primers (Liefers et al., 1998, New Engl J Med 339:223-8). Nested RT-PCR failed to yield CEA-specific amplicons in reactions using total RNA from patients in the control group, but detected carcinoembryonic antigen-specific amplicons in 1 patient in the case group.

35

The presence of carcinoembryonic antigen-specific amplicons was confirmed by sequence analysis.

GCC mRNA expression in lymph nodes and clinicopathological prognostic indicators. Case and control groups (28 patients) were compared for

- 5 tumor and disease characteristics associated with disease recurrence. Groups appeared balanced with respect to: tumor grade (well differentiated: control, 2 (12.5%); case, 1 (8.3%); moderately differentiated: control, 13 (81.3%); case, 9 (75%); poorly differentiated: control, 1 (8.3%); case, 2 (12.5%); tumor size (control, 5.7 ± 2.3 cm; case, 4.8 ± 1.7 cm); tumor location (right colon: control, 7 (43.8%);
- 10 case, 4 (33.3%); transverse colon: control, 3 (18.8%); case, 0; sigmoid colon: control, 5 (31.3%); case, 8 (66.6%); rectum: control, 1 (6.3%), case, 0); and depth of penetration and extension into pericolic fat of tumors. Angiolymphatic invasion was observed in 3 patients in the case group but not in patients in the control group, reflecting a likely mechanism underlying metastasis in the former. Expression of
- 15 GCC mRNA in lymph nodes was associated with disease recurrence in all cases ($p=0.004$). The odds ratio for mortality associated with GCC mRNA expression in regional lymph nodes was 16.5 (1.1 – 756.7, 95% CI). Sensitivity analysis demonstrated that an incremental “false negative” (death of a patient in the control group) or “false positive” (survival of a patient in the case group) result would yield
- 20 an odds ration with a 95% confidence interval encompassing 1 (no excess risk), reflecting the limitations of the small sample population employed in this analysis.